Chapter 6

Quantification of age-related structural genome changes in blood cells using whole genome sequencing

Tristan V. de Jong, Marianna R. Bevova, Szymon M. Kielbasa, Genome of the Netherlands Consortium, Kai Ye, Peter M. Lansdorp, Victor Guryev

Manuscript in preparation

ABSTRACT

Genetic alterations are one of the key hallmarks of ageing. Aneuploidy, shortening of the telomeres, accumulation of somatic mutations and activation of the transposable elements are examples of reported age-related changes. Rapidly increasing whole genome sequencing (WGS) data offer a practical way to generate accurate profiles of DNA alterations in individual genomes with age. To explore the utility of WGS data for quantification of age-related structural genome changes, we investigated WGS data obtained from blood cells of 767 individuals aged between 19 and 87 years.

We replicated previous studies reporting significant correlations of age to telomere length, mitochondrial DNA (mtDNA) content, loss of sex chromosomes, T-cell depletion and activation of LINE1 retrotransposition. Most of these changes fit a linear decline, except for the sporadic loss of chromosome Y and activation of LINE1 elements which both accelerate with age. Comparison of age-corrected values exposed a strong correlation between decreased mtDNA copy numbers and T-cell depletion, supporting a role of mitochondrial metabolism in T-cell exhaustion. Intriguingly, we found that while loss of the Y chromosome corresponds to shorter telomeres, loss of an X chromosome in females correlated with longer telomeres, highlighting the complexity in the relation between telomeres and aneuploidy. Several of samples showed a substantially elevated rate of LINE1 transposition, with the largest contribution coming from evolutionary younger elements. In those individuals we identified the X chromosome as a primary source of transposing LINEs. These findings and our bioinformatics tools and approach should facilitate further studies on genome ageing using WGS data.

MAIN TEXT

One of the hallmarks of ageing is the accumulation of changes in genetic material[1]. Increased aneuploidy[2], point mutations[3], shortening of the telomeres[4], depletion of mitochondrial DNA (mtDNA)[4] and activation of retrotransposons[5] are known examples of age-related changes in our genomes. The technical advances in genome

sequencing technologies have inspired large whole-genome sequencing (WGS) projects and have enabled studies of age-related changes in genome-wide manner [6,7]. In this study we used the sequencing data from the Genome of the Netherlands project8 to investigate the range of age-related structural genome changes in human blood cells (Fig 1).

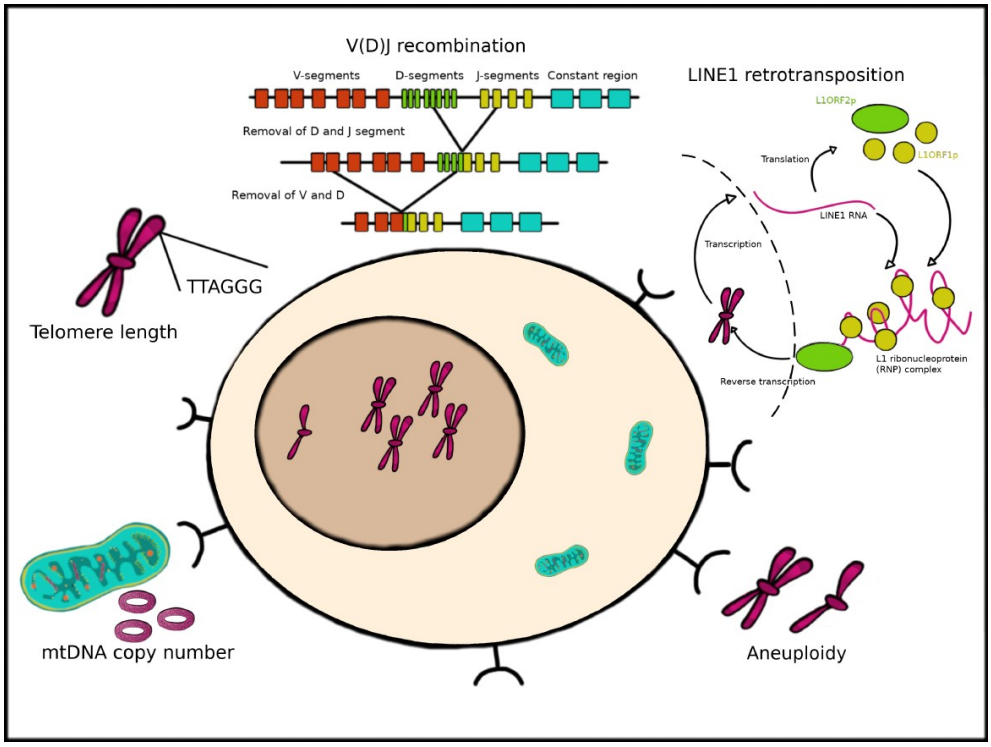

Figure 1. Known age-related changes in human DNA and outline of the experiment. With age, Telomere length decreases, mitochondrial content decreases, X and Y chromosome loss occurs, LINE1 retrotransposition events accumulate and the proportion of T-cells in blood goes down, which can be monitored by amount of V(D)J rearranged TCR loci.

Telomere shortening. To validate our WGS-based approach of identifying age-related genome changes we quantified proportion of reads containing multiple instances of telomeric DNA motif (GGGTTA). A significant correlation with age in gender-dependent manner was observed, in line with previous reports [7,9]. The telomere loss explains about 17.2% and 11.2% of the variance of male and female ages,

respectively (Fig. 2a, Table 1). Males exhibit a significantly higher rate of telomere loss (Table 1).

mtDNA depletion. Next, we quantified the relative abundance of reads mapping to mitochondrial DNA and its correlation with the age of individuals. Previous studies used WGS data for mtDNA copy-number estimation and reported a linear decline of mtDNA copy number in blood lymphocytes from 2000 Sardinian samples[11]. Our data suggest a linear decline of mtDNA copy number in males, while no significant decrease was observed in females (Fig 2b).

TABLE 1. Age-dependent molecular characteristics detected in genome sequencing.

Hallmark	Gender	Significance DNA isolation	Age	Base level at 20 years old***	Change per year
Telomere length	2.4×10^{-3}	0.86	$<10^{-5}$	F: 341.2 kb M: 341.6 kb	-1.68 kb -1.94 kb
mtDNA content	$<10^{-5}$	1.0×10^{-5}	8.3×10^{-4}	P: 166.9 FPM Q: 705.1 FPM	nonlinear
Chromosome X	Females only	8.6×10^{-5}	$<10^{-5}$	P: 60,553 FPM Q: 57,619 FPM	-11.8 FPM -16.9 FPM
Chromosome Y*	Males only	0.01	$<10^{-5}$ **	P: 2,084 FPM Q: 2,013 FPM	sporadic loss
T-Cell proportion	$<10^{-5}$	8.5×10^{-5}	$<10^{-5}$	P: 0.029 FPM Q: 0.053 FPM	-1.5×10^{-4} FPM -2.9×10^{-4} FPM
LINE/L1 activity	0.02	0.46	$<10^{-5}$ **	858.3 FPM	4.13 FPM

* Only reads mapping to male-specific part of Chromosome Y were quantified
** Rank test was used to estimate p-value because of apparent non-linear dependency of parameter with age
*** F - female, M - male, P- phenol/chloroform DNA isolation, Q- Qiagen DNA isolation

Effect of DNA isolation method. It is known that DNA isolation methods play an important role in the representation of different genome segments, and data analysis should account for this experimental variation[12]. We noticed a large effect of DNA isolation method on the mtDNA copy number and three other structural signatures tested in this study (Table 1; Supplementary Fig. 1). Since DNA isolation for majority of samples in this study (n=603) was performed with Qiagen kits, we excluded samples isolated through phenol-chloroform extraction from the analysis of the four signatures significantly affected by DNA isolation procedure. The estimates on

telomere length and LINE1 retro-transposition (discussed below), on the other hand, did not show significant differences and were studied for all 767 individuals.

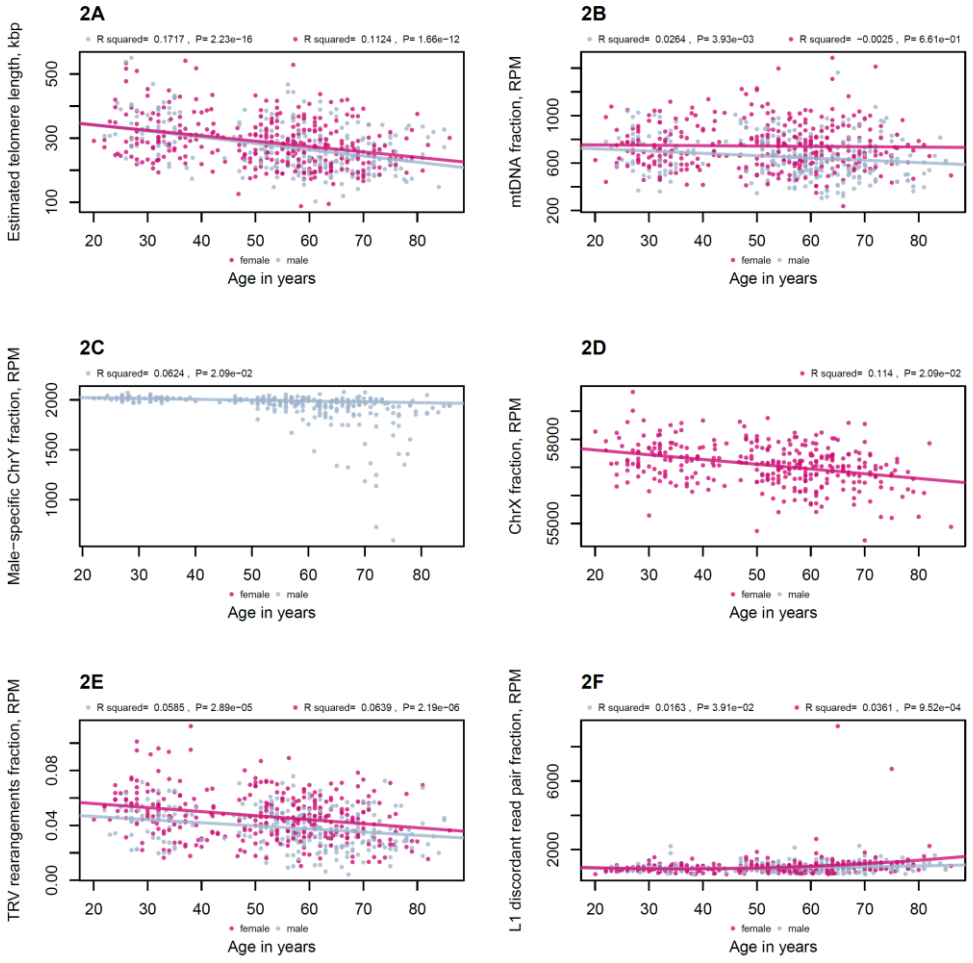

Figure 2. Correlation between age of individuals and genome characteristics **(A)** Estimated telomere sizes (kilobase per telomere); **(B)** mtDNA read content, reads per million (RPM) **(C)** Read content (RPM) corresponding to male-specific part of Y chromosome. Only male samples are shown; **(D)** Proportion of reads (RPM) mapping to X chromosome; Only female samples are shown; **(E)** Proportion of read pairs (RPM) compatible with rearranged T-cell receptor segments; **(F)** Estimated read content (RPM) of discordantly mapped read pairs where one of the reads corresponds to LINE/L1 repeat.

Loss of sex chromosomes. The read content that corresponds to the male-specific part of chromosome Y (MSY) exhibited a sporadic decrease in males (Fig. 2c). The read loss is consistent across different loci within MSY and thus likely represents complete loss of the Y chromosome events in a sizeable fraction of blood cells[2,13]. In females, we observed an almost linear age-dependent loss of X chromosomes (Fig. 2d). The loss of inactivated X chromosome was reported in several previous studies[14–17]. Using linear regression, we estimated that about 3.3% of cells have lost an X chromosome in females between 20 and 80 years of age. No significant age-related change of X chromosomes was detected in males. In contrast to sex chromosomes the read content of the autosomes did not show significant changes with age (Supplementary table S1).

Depletion of T-cells. High genome coverage of population-based genome sequencing project allows quantification of regional changes in chromosome structure. We observed a significant decrease of discordantly mapped read pairs corresponding to genes in TCR loci (Figure 2e, Supplementary Table S2). The quantified fragments correspond to V(D)J-rearranged versions of T-cell receptor genes that are characteristic for mature T-cells[18]. The observed decrease could be linked to the age-related thymus involution[19–21], the relative decrease of T-cell counts among the blood cells[22] and the observation that telomere loss is most pronounced in memory T-cells[10].

Increase of LINE1 retro-transposition events. A proportion of discordantly mapped read pairs that overlap with annotated LINE1 retro-transposable elements showed a significant increase (Figure 2f, Supplementary Table 3) with age. This finding is in line with the previously observed increase of LINE1 copy number with age in mouse tissues[23]. Interestingly, next to the generally pronounced age-dependent increase of discordant LINE1 reads, several samples exhibit a very high content of discordant read pairs that involve LINE1 sequences. The top 5 samples with the highest content of discordantly mapped pairs overlapping with LINE1 elements (over 2.8-fold increase over median value in our cohort) are females aged 61 and older. The striking increase was consistent across independently prepared and sequenced WGS libraries

and was not seen in their progeny (Fig. 3a), indicating somatic activation of LINE1 elements. Further studies will show if a gender bias in sporadic LINE1 activation is significant and whether such activation has measurable effects on health.

We checked if an age-related increase of LINE1 elements is associated with particular element(s) or chromosomal locations (Figure 3b). Evolutionary younger primate-specific elements (L1P) showed increase in mobility compared to other families of LINE1 elements. Similarly, part of mammalian-specific elements (L1MA) exhibited somatic expansion with a higher rate compared to other subfamilies (L1MB - L1ME, Figure 3b).

We explored how potential LINE1 retro-transposition events are distributed throughout the human genome. We counted robustly mapping reads that overlap an annotated LINE1 element (L1 source) and have paired read mapping to different chromosome not overlapping with known LINE1 (L1 destination). Our comparison of the density of mapped reads that form discordant pair with read containing LINE1 sequence showed more frequent transposition from the X chromosome (Figure 3 bottom). Previous studies reported that the X chromosome contains a higher proportion of LINE1 elements compared to autosomes. This 2-fold enrichment in LINE1 elements might be important for X chromosome inactivation[24]. We suggest that, next to a potential selective advantage for LINE1 accumulation on the X chromosome, LINE1 elements on that chromosome might be transcriptionally more active, possibly as a result of X-specific gene dosage compensation. We did not find global hotspots of LINE1 destinations, indicating a random integration along human chromosomes.

Dynamics of age-related genome changes. We performed linear and quadratic fittings between six structural features and age of individuals (Supplementary figure 4). The estimated speed of LINE1 retro-transposition increases with age and is best described by a quadratic curve (Fig 2f, inset). The loss of Y chromosomes shows a sporadic pattern. It was not observed before the age of 50, while the proportion of cells that lost this chromosome is highly variable in elderly males (Fig 2d). The other four

of the genomic changes are best described by linear declines (Fig 1a-c,e). This shows that, in blood cells, somatic genome alterations continue to accumulate during the whole lifespan.

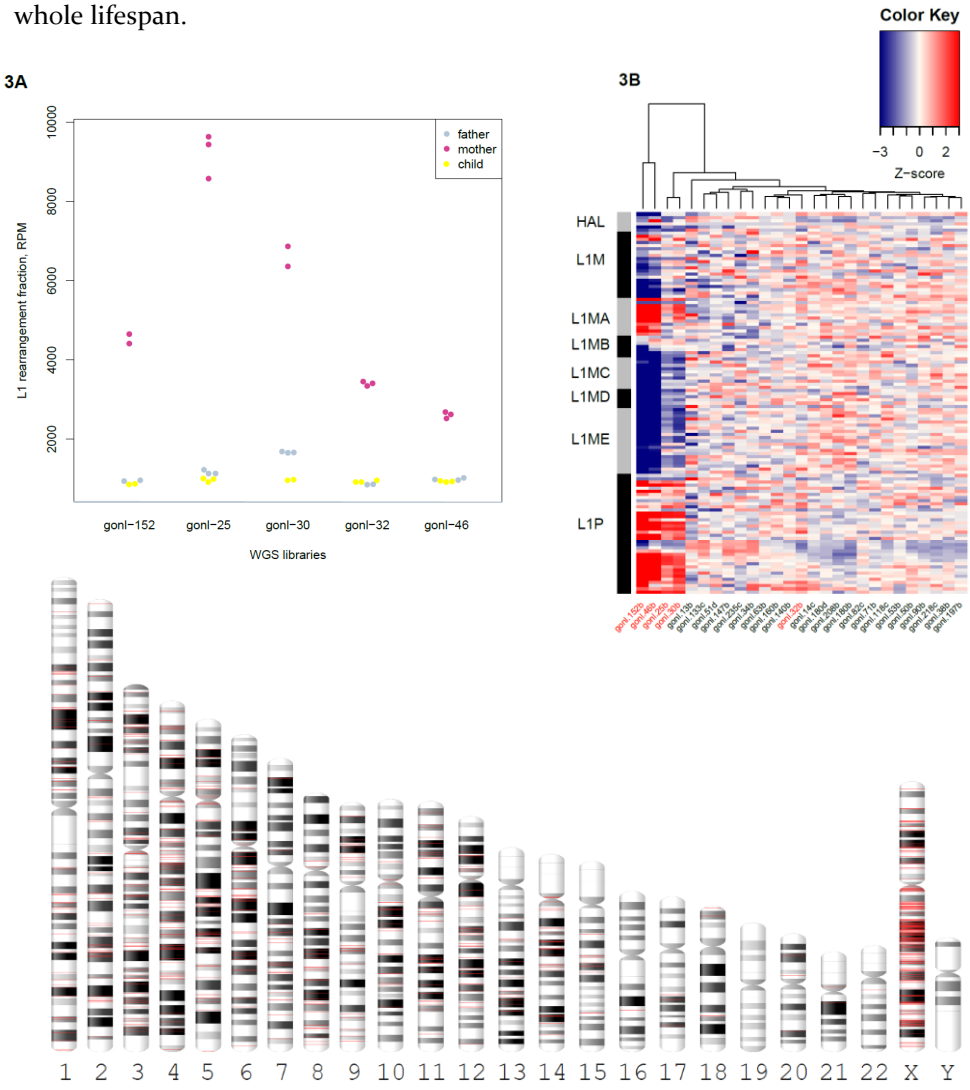

Figure 3. Somatic LINE1 mobility events are represented unevenly among repeat elements and along chromosomes **(A)** Estimated LINE1 activity (proportion of discordant read pairs overlapping with LINE1 elements, RPM) per library in five families where mother showed significant increase in LINE1 mobility; Note that the changes are consistent between different libraries of the same individual and offspring so not exhibit LINE1 mobility increase, indicating somatic source of this variants **(B).** Heatmap of the L1 outlier samples **(in red)** with 22 random samples **(in black)**. Primate-specific (L1P) and mammalian-specific (L1MA) elements show larger increase in number of discordant pairs compared to other L1 elements. **(bottom)** Chromosome distribution of L1 source regions shows relatively frequent activation of LINE1 elements on chromosome X.

The variance in the age groups may reflect difference in lifestyle, environment and disease status. Future studies incorporating larger cohorts and different tissues, might reveal more subtle genomic changes that are characteristics for ageing process.

Connection among ageing signatures. Next, we investigated if telomere attrition, loss of mtDNA, loss of sex chromosomes, depletion of T-cells and activation of LINE1 retro-transposition are inter-related. We expressed the observed values relative to the values expected for the given age. Checking correlations between different features allowed us, for example, to test if individuals having a low copy-number of mtDNA expected for their age also showed shorter telomere size compared to their peers.

We observed four significant relations ($p < 0.01$) between aging signatures (supplementary figure 2A, B). Two positive correlations, though small, link mtDNA copy number with T-cell depletion and Telomere length (two tailed T-Test; $p=1.13e^{-9}$; $p=8.87e^{-9}$ Fig. 4 A-B). Interestingly, recent reports did indicate a possible relation between defects in mitochondrial metabolism and T-cell exhaustion[25,26]. The other two significant negative correlations connected loss of chromosome X In females with telomere length and T-cell depletion (Fig 4 C, D; two-tailed T-Test; $p=3.6e^{-19}$ and $p=5.2e^{-4}$ respectively).

The former anti-correlation is surprisingly strong with age-corrected telomere length explaining 21.6% of variance in relative loss of chromosome X. Previously, chromosomal loss has been shown to be positively associated with the chromosome's telomere length in human somatic cells[27]. Our findings of a negative relation indicate that telomere shortening can play a dual role in somatic aneuploidy.

Scan for other age-related genome features. Using window-based whole-genome scan we tested other genomic footprints that can be associated with age (Supplementary table S2). This search identified several regions characterized by age-specific loss of concordantly mapped reads pairs (telomeric regions), discordant reads mapping to the same chromosome (TCR locus on Chr14) and discordant reads mapping to different chromosomes (loci with full-length LINE1 elements).

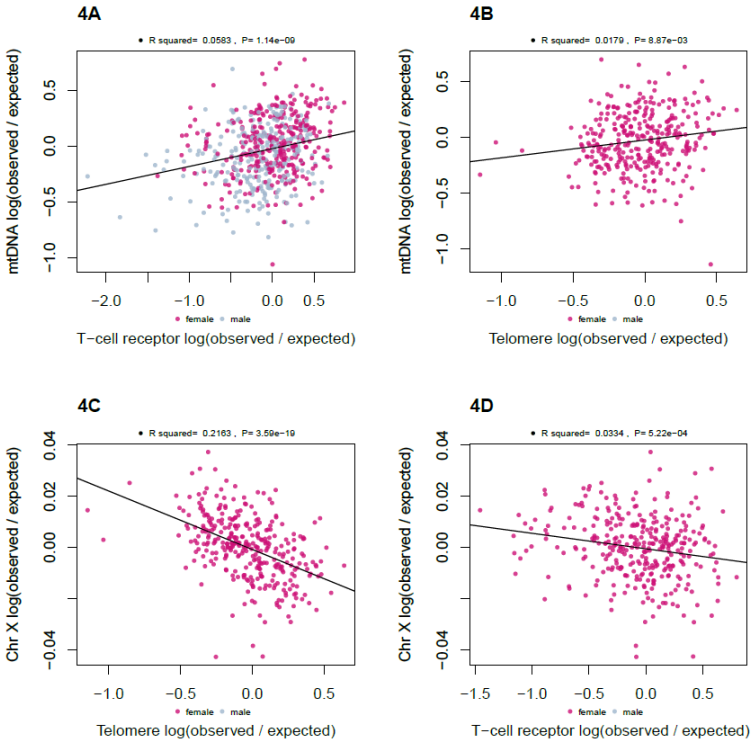

Figure 4. Correlation between age-corrected molecular signatures of ageing. Log2-ratio of observed values and values expected for reported individual age are plotted. **(A-D)** Negative correlations between T-cell receptor fraction and mtDNA content **(A),** Telomere length and mtDNA content **(B),** chr X and Telomere length **(C)**, T-cell receptor fraction and chr X content **(D).**

Genome partitioning based on repeat and gene types (Supplementary table S3) showed an increase in the number of discordant pairs mapping to different chromosomes where one of the reads overlaps a various LINE1 elements (mostly from L1P and L1MA subfamilies). These results indicate that the six genomic hallmarks of ageing studied here represent a set of signatures with the largest age-specific effects. Future studies employing larger cohorts might reveal more subtle age-dependent changes in genome structure.

Tool for quantification of age-related genome changes from WGS data. To assist these future studies, we developed a tool, DNAgeQuant that allows to quantify the structural genome changes and overlay them with results obtained for GoNL cohort.

The tool (http://tools.genomes.nl/DNAgeQuant) processes paired-end data aligned to a recent version of human genome reference (e.g. GRCh37 or GRCh38) and calculates factors such as telomere length, content of mtDNA, Chromosome X, male-specific part of chromosome Y, proportions of discordant pairs for T-Cell receptor loci and LINE1 elements. It can also accept custom annotation representing genomic features of interest to test potential new signatures associated with disease or ageing. The program can visualize the results in the context of GoNL data (supplementary fig 3). When age of subject is specified, program will present values relative to expected for the given age (expressed as Z-scores or percent, Figure 5).

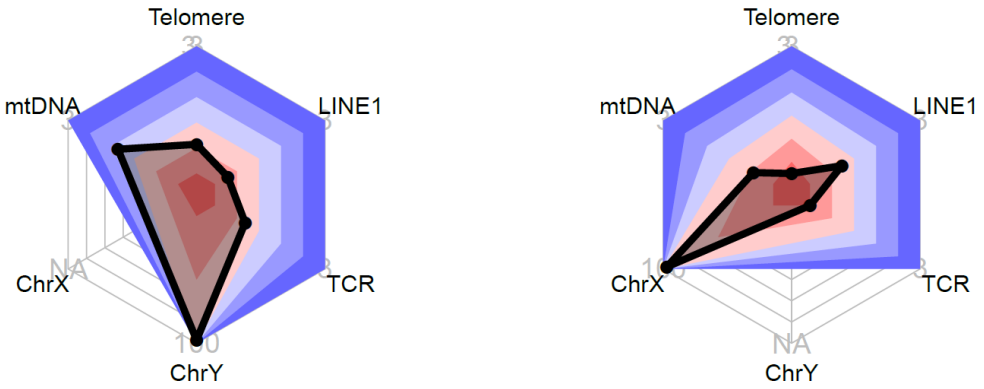

Figure 5. Visualization of age-related structural genome changes quantified with DNAgeQuant software. A 58-year-old male is shown on the left, whilst a 58 year old female is shown on the left. The decline of the several factors is shown in Z-scores as compared to their peers, or a percentage of the average total of 20-year-old samples. The divide between red and blue represents a deviation of 0 SD, each zone outward means 1 SD increase (blue) or decrease (red) from the expected value for the age of the sample.

Estimating age from structural genome features. To explore if the variety of molecular quantifications might have utility for estimating age of individual (e.g. for forensic studies), we tested various combination of features for age prediction. The model that combines telomere length, chromosome X, TCR and LINE1 measures in females can explain 37% of age variation (SD=11.7 years). The model that includes telomere length, mtDNA, chromosome Y and LINE1 explained 30% of variance in male

age (SD=12.7 years; Supplementary Figure 4). In both sexes telomere length and the corresponding sex chromosome loss lend the biggest contribution for the prediction of age.

CONCLUSIONS

Taken together, our results show that the human genome undergoes many structural changes that are correlated with age. Most of the changes observed in this study show a linear correlation with age, indicating continuous accumulation of structural genome changes throughout the whole lifespan. The relations among the features seem complex but taken together they can improve our ability to predict age and advance our understanding of genetic and environmental factors involved in genome alterations and ageing. The automated profiling of these DNA alterations in individual genomes using whole genome sequencing data opens new avenues for the interpretation of genome data.

METHODS

Sample selection, genome sequencing and alignment. Data from GoNL consortium were used for our analysis[28]. For this cohort 769 individuals (354 males and 411 females) of self-reported Dutch ancestry have been selected. Five Dutch biobanks from different parts of the country have contributed samples for the GoNL project. a) LifeLines cohort study (n=165); b) The Leiden Longevity Study (n=72); c) The Netherlands Twin Registry (n=349); d) Rotterdam Study (n=102); and e) Erasmus Rucphen Family Study (ERF; n=81). The sampling has not been based on a specific phenotype or disease. The cohort consists of 231 trios, 11 quartets with monozygotic twins and 8 quartets with dizygotic twins. The parent-offspring trios include adult individuals in age range from 19 to 87 years (mean = 53 years, SD= 16 years. More details on cohort and the principles of sample selection were described earlier[28]. The DNA has been extracted from blood samples. Whole genome sequencing of all samples to a median coverage of 13.5x was done on Illumina HiSeq 2000 sequencers

(90-bp paired-end reads, ~500-bp insert size). Quality filtering and read alignment were described previously[8]. From the current analysis we have excluded two individuals for which our sequencing data have indicated XXX and XXY karyotypes.

Estimation of telomere length from the sequencing data. Telomeres length have been estimated from the whole-genome sequence data using previously published method[7]. The telomeric reads have been defined as reads containing 7 or more repeats TTAGGG. These later have been translated into an estimate of the physical length of telomeres per haploid genome using total number of concordantly mapped reads mapped to autosomes (Chr1 to Chr22) and total length of autosomes. Sex chromosomes and mtDNA were excluded from normalization as they exhibited age-dependent variation in their copy-numbers.

Estimation of chromosome/mtDNA loss. For each individual we counted the number of concordantly mapped read pairs per chromosomes X and MT. Quantification of Y chromosome segments is known to be complicated by segments with high homology to ChrX. We, therefore, only counted reads mapped as concordant pairs to male-specific part of ChrY (MSY, bases 6,611,498 to 24,510,581) required unambiguous mapping (mapping quality >= 20, estimated 1% of mapping error) and excluded seven other small regions within MSY that showed read coverage in female samples, see DNAgeQuant program for detailed list of such regions. Number of reads mapping to X, MSY and MT regions was normalized using total number of reads concordantly mapped to autosomes.

Estimation of T-cell proportion. We quantified the proportion of read pairs that represent V(D)J rearranged segments that involve TRV genes as a proxy for proportion of T-cells in a sample. The location of TRV genes were fetched from Ensembl release 75. We counted reads in discordantly mapped pairs where one read overlapped a TRV gene and its pair was located on the same chromosome. The number or reads corresponding to rearranged T-cell receptor genes was normalized by total number of reads mapped to autosomes.

Estimation of LINE1 retrotransposion activity. We quantified the proportion of read pairs that indicate LINE1 element retrotransposion events as a proxy for LINE1 activity in a sample. The genomic locations of LINE1 elements were fetched from Ensembl release 75. We counted reads in discordantly mapped pairs where one read overlapped a LINE1 element and its pair was mapped on the different chromosome. We argued that discordant pairs within a single chromosome can also be a result of other types of structural variants, such as deletions, duplications or inversions. The read number corresponding to potential LINE1 retrotranspositions was normalized by total number of reads mapped to autosomes.

Testing of effects of sex and DNA isolation. The analysis was done using LME4 package[29]. Random intercept models accounting for age, sex and DNA isolation method (fixed effects), sequencing batch and family (random effects) were used. Data was checked for linearity and normality of residuals. Significance was determined using likelihood ratio test.

Analysis of LINE1 active elements and hotspots. We counted robustly mapping reads (mapping quality ≥20) that overlap an annotated LINE1 element (source) and have paired read mapping to different chromosome not overlapping with known LINE1 (destination). Per element values for each sample were expressed as fractions of discordant LINE1 read pairs. For each element we calculated mean and standard deviation across all samples and used Z-scores to determine LINE1 elements showing increased retrotransposition rate in 5 samples with highest level of LINE1 transposition. Similar analysis was performed in 100kb-long genomic windows for source and destination locations of LINE1 retrotransposition events. Genomic regions that showed z-score of 3 and higher were considered as hotspots.

Fitting of ageing signatures to linear and quadratic models. Fitting of the data to linear (e.g. Telomere ~ Age) and quadratic (e.g. LINE1_activity ~ Age + Age^2) using *lm* function implemented in R version 3.2.1.

Correlations between age-corrected values between features. We expressed the observed values relative to the values expected for the given age (Y chromosome content for individuals aged 19-50 years for ChrY, quadratic regression for LINE1 retrotranspositions and linear regressions for the rest of the features). Log2 transformed ratios of observed versus expected (for this age, based on regression) were used for analysis using Pearson correlation.

Estimation of gene/repeat dynamics. We performed scans using whole-chromosomes copy numbers, 100 kilobase-long genomic windows, overlap with repeat elements (classes, types and individual elements), overlaps with individual genes, and gene biotypes. We repeated analysis for several types or read pairs: concordantly mapped, discordantly mapped within same chromosome, discordantly mapped to different chromosomes. Repeat annotation for GRCh37 assembly was downloaded from http://www.ensembl.org, release 75 to identify the genomic coordinates of annotated genes, repetitive elements and low-complexity sequences. Using DNAgeQuant program (http://tools.genomes.nl/DNAgeQuant) for each sample we obtained raw counts of reads belonging to each of 3 type of pairs (concordantly mapped, discordantly mapped within same chromosome, discordantly mapped to different chromosomes) overlapping different genomic features (chromosomal regions, genes, gene biotypes, repeat element, repeat class, repeat type).

The data were analyzed using the edgeR package[30]. In each case we fitted a generalized linear model to raw counts of read pairs for each individual and each genome window, repeat or gene type. The model considers effects of sex, biobank, sequencing batch, and proportion of reads mapped to male-specific part of chromosome Y (to account for loss of Y-chromosome). Nominal p-value was determined by likelihood ratio test. Multiple testing correction with Benjamini-Hochberg method was used to determine false discovery rate estimates.

Tool for quantification of structural changes in genome. We developed DNAgeQuant tool for automatic quantification of age-related genome features. The tool is implemented as Perl script, takes in SAM/BAM/CRAM file as input and depends on

Samtools program[31]. The tool generates several R scripts for visualization of plots and age estimation. The computational performance of the tool is largely affected by BAM/CRAM decompression.

Age prediction from WGS data. Multiple regression model combining 4 factors for females (telomere length, X chromosome, T-cell receptor and LINE1 retrotransposition) or 5 factors for males were used to predict age from structural genome features.

REFERENCES

1. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153,** 1194–217 (2013).

2. Forsberg, L. A. *et al.* Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat. Genet.* **46,** 624–8 (2014).

3. Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538,** 260–264 (2016).

4. Blackburn, E. H., Epel, E. S. & Lin, J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science* **350,** 1193–8 (2015).

5. Gorbunova, V., Boeke, J. D., Helfand, S. L. & Sedivy, J. M. Human Genomics. Sleeping dogs of the genome. *Science* **346,** 1187–8 (2014).

6. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

7. Ding, Z. *et al.* Estimating telomere length from whole genome sequence data. *Nucleic Acids Res.* **42,** e75 (2014).

8. Francioli, L. C. *et al.* Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46,** 818–825 (2014).

9. Gardner, M. *et al.* Gender and telomere length: systematic review and meta-analysis. *Exp. Gerontol.* **51,** 15–27 (2014).

10. Aubert, G., Baerlocher, G. M., Vulto, I., Poon, S. S. & Lansdorp, P. M. Collapse of telomere homeostasis in hematopoietic cells caused by heterozygous mutations in telomerase genes. *PLoS Genet.* **8,** e1002696 (2012).

11. Ding, J. *et al.* Assessing Mitochondrial DNA Variation and Copy Number in Lymphocytes of ~2,000 Sardinians Using Tailored Sequencing Analysis Tools. *PLoS Genet.* **11,** e1005306 (2015).

12. van Heesch, S. *et al.* Systematic biases in DNA copy number originate from isolation procedures. *Genome Biol.* R33 (2013). doi:10.1186/gb-2013-14-4-r33

13. Dumanski, J. P. *et al.* Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease. *Am. J. Hum. Genet.* **98,** 1208–19 (2016).

14. Guttenbach, M., Koschorz, B., Bernthaler, U., Grimm, T. & Schmid, M. Sex chromosome loss and aging: in situ hybridization studies on human interphase nuclei. *Am. J. Hum. Genet.* **57,** 1143–50 (1995).

15. Surrallés, J., Hande, M. P., Marcos, R. & Lansdorp, P. M. Accelerated Telomere Shortening in the Human Inactive X Chromosome. *Am. J. Hum. Genet.* **65,** 1617–1622 (1999).

16. Russell, L. M., Strike, P., Browne, C. E. & Jacobs, P. A. X chromosome loss and ageing. *Cytogenet. Genome Res.* **116,** 181–5 (2007).

17. Tang, Z. *et al.* Clinical significance of acquired loss of the X chromosome in bone marrow. *Leuk. Res.* **47,** 109–13 (2016).

18. Johnson, P. L. F., Goronzy, J. J. & Antia, R. A population biological approach to understanding the maintenance and loss of the T-cell repertoire during aging. *Immunology* **142,** 167–175 (2014).

19. Haynes, B. F., Sempowski, G. D., Wells, A. F. & Hale, L. P. The human thymus during aging. *Immunol. Res.* **22,** 253–61 (2000).

20. Taub, D. D. & Longo, D. L. Insights into thymic aging and regeneration. *Immunol. Rev.* **205,** 72–93 (2005).

21. Ki, S. *et al.* Global transcriptional profiling reveals distinct functions of thymic stromal subsets and age-related changes during thymic involution. *Cell Rep.* **9,** 402–15 (2014).

22. Qin, L. *et al.* Aging of immune system: Immune signature from peripheral blood lymphocyte subsets in 1068 healthy adults. *Aging (Albany. NY).* **8,** 848–59 (2016).

23. De Cecco, M. *et al.* Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging (Albany. NY).* **5,** 867–83 (2013).

24. Bailey, J. A., Carrel, L., Chakravarti, A. & Eichler, E. E. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6634–9 (2000).

25. Bengsch, B. *et al.* Bioenergetic Insufficiencies Due to Metabolic Alterations Regulated by the Inhibitory Receptor PD-1 Are an Early Driver of CD8(+) T Cell Exhaustion. *Immunity* **45,** 358–73 (2016).

26. Scharping, N. E. *et al.* The Tumor Microenvironment Represses T Cell Mitochondrial Biogenesis to Drive Intratumoral T Cell Metabolic Insufficiency and Dysfunction. *Immunity* **45,** 374–388 (2016).

27. Van Leeuwen, E. M. *et al.* Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. *Nat. Commun.* **6,** (2015).

28. Boomsma, D. I. *et al.* The Genome of the Netherlands: Design, and project goals. *Eur. J. Hum. Genet.* **22,** 221–227 (2014).

29. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67,** 1–48 (2015).

30. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–40 (2010).

31. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–9 (2009).

SUPPLEMENTARY MATERIAL

In order to let the reader to look at really small points on really large plots, all supplemental material of this thesis have been moved online. For additional reading, please see the following link:

https://drive.google.com/open?id=1skLP9E2hToDLGgkNrmCoWcTVeBmiZX1-

If you might encounter a broken link, please contact me:

Tristan_dejong@hotmail.com